# Bifix codes and interval exchanges

Valérie Berthé, Clelia de Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, Giuseppina Rindone

# Bifix codes and interval exchanges

Valérie Berthé[1], Clelia De Felice[2], Francesco Dolce[3], Julien Leroy[4], Dominique Perrin[3], Christophe Reutenauer[5], Giuseppina Rindone[3]

[1]CNRS, Université Paris 7, [2]Università degli Studi di Salerno, [3]Université Paris Est, LIGM, [4]Université du Luxembourg, [5]Université du Québec à Montréal

July 22, 2014 18 h 45

**Abstract**

We investigate the relation between bifix codes and interval exchange transformations. We prove that the class of natural codings of regular interval echange transformations is closed under maximal bifix decoding.

## Contents

**1 Introduction** **2**

**2 Interval exchange transformations** **3**
- 2.1 Regular interval exchange transformations . . . . . . . . . . . . . 4
- 2.2 Natural coding . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
- 2.3 Uniformly recurrent sets . . . . . . . . . . . . . . . . . . . . . . 6
- 2.4 Interval exchange sets . . . . . . . . . . . . . . . . . . . . . . . 7

**3 Bifix codes and interval exchange** **9**
- 3.1 Prefix codes and bifix codes . . . . . . . . . . . . . . . . . . . . 9
- 3.2 Maximal bifix codes . . . . . . . . . . . . . . . . . . . . . . . . 10
- 3.3 Bifix codes and regular transformations . . . . . . . . . . . . . . 13

**4 Tree sets** **16**
- 4.1 Extension graphs . . . . . . . . . . . . . . . . . . . . . . . . . . 16
- 4.2 Interval exchange sets and planar tree sets . . . . . . . . . . . . . 17
- 4.3 Exchange of pieces . . . . . . . . . . . . . . . . . . . . . . . . . 18
- 4.4 Subgroups of finite index . . . . . . . . . . . . . . . . . . . . . . 21

# 1 Introduction

This paper is part of a research initiated in [2] which studies the connections between the three subjects formed by symbolic dynamics, the theory of codes and combinatorial group theory. The initial focus was placed on the classical case of Sturmian systems and progressively extended to more general cases.

The starting point of the present research is the observation that the family of Sturmian sets is not closed under decoding by a maximal bifix code, even in the more simple case of the code formed of all words of fixed length $n$. Actually, the decoding of the Fibonacci word (which corresponds to a rotation of angle $\alpha = (3 - \sqrt{5})/2$) by blocks of length $n$ is an interval exchange transformation corresponding to a rotation of angle $n\alpha$ coded on $n + 1$ intervals. This has lead us to consider the set of factors of interval exchange transformations, called interval exchange sets. Interval exchange transformations were introduced by Oseledec [15] following an earlier idea of Arnold [1]. These transformations form a generalization of rotations of the circle.

The main result in this paper is that the family of regular interval exchange sets is closed under decoding by a maximal bifix code (Theorem 3.13). This result invited us to try to extend to regular interval exchange transformations the results relating bifix codes and Sturmian words. This lead us to generalize in [5] to a large class of sets the main result of [2], namely the Finite Index Basis Theorem relating maximal bifix codes and bases of subgroups of finite index of the free group.

Theorem 3.13 reveals a close connection between maximal bifix codes and interval exchange transformations. Indeed, given an interval exchange transformation $T$ each maximal bifix code $X$ defines a new interval exchange transformation $T_X$. We show at the end of the paper, using the Finite Index Basis Theorem, that this transformation is actually an interval exchange transformation on a stack, as defined in [7] (see also [19]).

The paper is organized as follows.

In Section 2, we recall some notions concerning interval exchange transformations. We state the result of Keane [12] which proves that regularity is a sufficient condition for the minimality of such a transformation (Theorem 2.3).

We study in Section 3 the relation between interval exchange transformations and bifix codes. We prove that the transformation associated with a finite $S$-maximal bifix code is an interval exchange transformation (Proposition 3.8). We also prove a result concerning the regularity of this transformation (Theorem 3.12).

We discuss the relation with bifix codes and we show that the class of regular interval exchange sets is closed under decoding by a maximal bifix code, that is, under inverse images by coding morphisms of finite maximal bifix codes (Theorem 3.13).

In Section 4 we introduce tree sets and planar tree sets. We show, reformulating a theorem of [9], that uniformly recurrent planar tree sets are the regular interval exchange sets (Theorem 4.3). We show in another paper [4] that, in the same way as regular interval exchange sets, the class of uniformly recurrent tree sets is closed under maximal bifix decoding.

In Section 4.3, we explore a new direction, extending the results of this paper to a more general case. We introduce exchange of pieces, a notable example being given by the Rauzy fractal. We indicate how the decoding of the natural codings of exchange of pieces by maximal bifix codes are again natural codings of exchange of pieces. We finally give in Section 4.4 an alternative proof of Theorem 3.13 using a skew product of a regular interval exchange transformation with a finite permutation group.

**Acknowledgements** This work was supported by grants from Région Ile-de-France, the ANR projects Eqinocs and Dyna3S, the Labex Bezout, the FARB Project "Aspetti algebrici e computazionali nella teoria dei codici, degli automi e dei linguaggi formali" (University of Salerno, 2013) and the MIUR PRIN 2010-2011 grant "Automata and Formal Languages: Mathematical and Applicative Aspects". We thank the referee for his useful remarks on the first version of the paper which lead us to separate it in two parts.

# 2 Interval exchange transformations

Let us recall the definition of an interval exchange transformation (see [8] or [6]).

A *semi-interval* is a nonempty subset of the real line of the form $[\alpha, \beta[ = \{z \in \mathbb{R} \mid \alpha \leq z < \beta\}$. Thus it is a left-closed and right-open interval. For two semi-intervals $\Delta, \Gamma$, we denote $\Delta < \Gamma$ if $x < y$ for any $x \in \Delta$ and $y \in \Gamma$.

Let $(A, <)$ be an ordered set. A partition $(I_a)_{a\in A}$ of $[0, 1[$ in semi-intervals is *ordered* if $a < b$ implies $I_a < I_b$.

Let $A$ be a finite set ordered by two total orders $<_1$ and $<_2$. Let $(I_a)_{a\in A}$ be a partition of $[0, 1[$ in semi-intervals ordered for $<_1$. Let $\lambda_a$ be the length of $I_a$. Let $\mu_a = \sum_{b\leq_1 a} \lambda_b$ and $\nu_a = \sum_{b\leq_2 a} \lambda_b$. Set $\alpha_a = \nu_a - \mu_a$. The *interval exchange transformation* relative to $(I_a)_{a\in A}$ is the map $T : [0, 1[ \to [0, 1[$ defined by

$$T(z) = z + \alpha_a \quad \text{if } z \in I_a.$$

Observe that the restriction of $T$ to $I_a$ is a translation onto $J_a = T(I_a)$, that $\mu_a$ is the right boundary of $I_a$ and that $\nu_a$ is the right boundary of $J_a$. We additionally denote by $\gamma_a$ the left boundary of $I_a$ and by $\delta_a$ the left boundary of $J_a$. Thus

$$I_a = [\gamma_a, \mu_a[, \quad J_a = [\delta_a, \nu_a[.$$

Note that $a <_2 b$ implies $\nu_a < \nu_b$ and thus $J_a < J_b$. This shows that the family $(J_a)_{a\in A}$ is a partition of $[0, 1[$ ordered for $<_2$. In particular, the transformation $T$ defines a bijection from $[0, 1[$ onto itself.

An interval exchange transformation relative to $(I_a)_{a\in A}$ is also said to be on the alphabet $A$. The values $(\alpha_a)_{a\in A}$ are called the *translation values* of the transformation $T$.

**Example 2.1** Let $R$ be the interval exchange transformation corresponding to $A = \{a, b\}$, $a <_1 b$, $b <_2 a$, $I_a = [0, 1-\alpha[$, $I_b = [1-\alpha, 1[$. The transformation $R$ is the rotation of angle $\alpha$ on the semi-interval $[0, 1[$ defined by $R(z) = z+\alpha \bmod 1$.

Since $<_1$ and $<_2$ are total orders, there exists a unique permutation $\pi$ of $A$ such that $a <_1 b$ if and only if $\pi(a) <_2 \pi(b)$. Conversely, $<_2$ is determined by $<_1$ and $\pi$ and $<_1$ is determined by $<_2$ and $\pi$. The permutation $\pi$ is said to be *associated* with $T$.

If we set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \cdots <_1 a_s$, the pair $(\lambda, \pi)$ formed by the family $\lambda = (\lambda_a)_{a\in A}$ and the permutation $\pi$ determines the map $T$. We will also denote $T$ as $T_{\lambda,\pi}$. The transformation $T$ is also said to be an $s$-interval exchange transformation.

It is easy to verify that if $T$ is an interval exchange transformation, then $T^n$ is also an interval exchange transformation for any $n \in \mathbb{Z}$.

**Example 2.2** A 3-interval exchange transformation is represented in Figure 3.2. One has $A = \{a, b, c\}$ with $a <_1 b <_1 c$ and $b <_2 c <_2 a$. The associated permutation is the cycle $\pi = (abc)$.

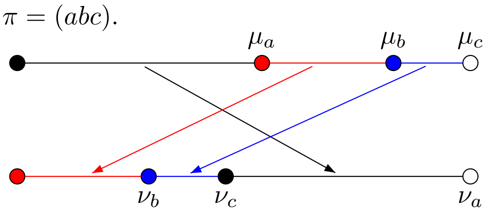


Figure 2.1: A 3-interval exchange transformation

## 2.1 Regular interval exchange transformations

The *orbit* of a point $z \in [0, 1[$ is the set $\{T^n(z) \mid n \in \mathbb{Z}\}$. The transformation $T$ is said to be *minimal* if, for any $z \in [0, 1[$, the orbit of $z$ is dense in $[0, 1[$.

Set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \ldots <_1 a_s$, $\mu_i = \mu_{a_i}$ and $\delta_i = \delta_{a_i}$. The points $0, \mu_1, \ldots, \mu_{s-1}$ form the set of *separation points* of $T$, denoted $\mathrm{Sep}(T)$. Note that the singular points of the transformation $T$ (that is the points $z \in [0, 1[$ at which $T$ is not continuous) are among the separation points but that the converse is not true in general (see Example 3.9).

An interval exchange transformation $T_{\lambda,\pi}$ is called *regular* if the orbits of the nonzero separation points $\mu_1, \ldots, \mu_{s-1}$ are infinite and disjoint. Note that the orbit of 0 cannot be disjoint of the others since one has $T(\mu_i) = 0$ for some $i$ with $1 \le i \le s-1$. The term regular was introduced by Rauzy in [17]. A regular interval exchange transformation is also said to be *without connections* or satisfying the *idoc* condition (where idoc stands for infinite disjoint orbit condition).

Note that since $\delta_2 = T(\mu_1), \ldots, \delta_s = T(\mu_{s-1})$, $T$ is regular if and only if the orbits of $\delta_2, \ldots, \delta_s$ are infinite and disjoint.

As an example, the 2-interval exchange transformation of Example 2.1 which is the rotation of angle $\alpha$ is regular if and only if $\alpha$ is irrational.

Note that if $T$ is a regular $s$-interval exchange transformation, then for any $n \geq 1$, the transformation $T^n$ is an $n(s-1)+1$-interval exchange transformation. Indeed, the points $T^i(\mu_j)$ for $0 \leq i \leq n-1$ and $1 \leq j \leq s-1$ are distinct and define a partition in $n(s-1)+1$ intervals.

The following result is due to Keane [12].

**Theorem 2.3 (Keane)** *A regular interval exchange transformation is minimal.*

The converse is not true. Indeed, consider the rotation of angle $\alpha$ with $\alpha$ irrational, as a 3-interval exchange transformation with $\lambda = (1-2\alpha, \alpha, \alpha)$ and $\pi = (132)$. The transformation is minimal as any rotation of irrational angle but it is not regular since $\mu_1 = 1-2\alpha$, $\mu_2 = 1-\alpha$ and thus $\mu_2 = T(\mu_1)$.

The following necessary condition for minimality of an interval exchange transformation is useful. A permutation $\pi$ of an ordered set $A$ is called *decomposable* if there exists an element $b \in A$ such that the set $B$ of elements strictly less than $b$ is nonempty and such that $\pi(B) = B$. Otherwise it is called *indecomposable*. If an interval exchange transformation $T = T_{\lambda,\pi}$ is minimal, the permutation $\pi$ is indecomposable. Indeed, if $B$ is a set as above, the set $S = \cup_{a \in B} I_a$ is closed under $T$ and strictly included in $[0,1[$.

The following example shows that the indecomposability of $\pi$ is not sufficient for $T$ to be minimal.

**Example 2.4** Let $A = \{a,b,c\}$ and $\lambda$ be such that $\lambda_a = \lambda_c$. Let $\pi$ be the transposition $(ac)$. Then $\pi$ is indecomposable but $T_{\lambda,\pi}$ is not minimal since it is the identity on $I_b$.

## 2.2 Natural coding

Let $A$ be a finite nonempty alphabet. All words considered below, unless stated explicitly, are supposed to be on the alphabet $A$. We denote by $A^*$ the set of all words on $A$. We denote by 1 or by $\varepsilon$ the empty word. We refer to [3] for the notions of prefix, suffix, factor of a word.

Let $T$ be an interval exchange transformation relative to $(I_a)_{a \in A}$. For a given real number $z \in [0,1[$, the *natural coding* of $T$ relative to $z$ is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots$ on the alphabet $A$ defined by

$$a_n = a \quad \text{if} \quad T^n(z) \in I_a.$$

For a word $w = b_0 b_1 \cdots b_{m-1}$, let $I_w$ be the set

$$I_w = I_{b_0} \cap T^{-1}(I_{b_1}) \cap \ldots \cap T^{-m+1}(I_{b_{m-1}}). \tag{2.1}$$

Note that each $I_w$ is a semi-interval. Indeed, this is true if $w$ is a letter. Next, assume that $I_w$ is a semi-interval. Then for any $a \in A$, $T(I_{aw}) = T(I_a) \cap I_w$ is a semi-interval since $T(I_a)$ is a semi-interval by definition of an interval exchange transformation. Since $I_{aw} \subset I_a$, $T(I_{aw})$ is a translate of $I_{aw}$, which is therefore also a semi-interval. This proves the property by induction on the length.

Set $J_w = T^m(I_w)$. Thus

$$J_w = T^m(I_{b_0}) \cap T^{m-1}(I_{b_1}) \cap \ldots \cap T(I_{b_{m-1}}). \tag{2.2}$$

In particular, we have $J_a = T(I_a)$ for $a \in A$. Note that each $J_w$ is a semi-interval. Indeed, this is true if $w$ is a letter. Next, for any $a \in A$, we have $T^{-1}(J_{wa}) = J_w \cap I_a$. This implies as above that $J_{wa}$ is a semi-interval and proves the property by induction. We set by convention $I_\varepsilon = J_\varepsilon = [0, 1[$. Then one has for any $n \geq 0$

$$a_n a_{n+1} \cdots a_{n+m-1} = w \Longleftrightarrow T^n(z) \in I_w \tag{2.3}$$

and

$$a_{n-m} a_{n-m+1} \cdots a_{n-1} = w \Longleftrightarrow T^n(z) \in J_w. \tag{2.4}$$

Let $(\alpha_a)_{a \in A}$ be the translation values of $T$. Note that for any word $w$,

$$J_w = I_w + \alpha_w \tag{2.5}$$

with $\alpha_w = \sum_{j=0}^{m-1} \alpha_{b_j}$ as one may verify by induction on $|w| = m$. Indeed it is true for $m = 1$. For $m \geq 2$, set $w = ua$ with $a = b_{m-1}$. One has $T^m(I_w) = T^{m-1}(I_w) + \alpha_a$ and $T^{m-1}(I_w) = I_w + \alpha_u$ by the induction hypothesis and the fact that $I_w$ is included in $I_u$. Thus $J_w = T^m(I_w) = I_w + \alpha_u + \alpha_a = I_w + \alpha_w$. Equation (2.5) shows in particular that the restriction of $T^{|w|}$ to $I_w$ is a translation.

## 2.3 Uniformly recurrent sets

A set $S$ of words on the alphabet $A$ is said to be *factorial* if it contains the factors of its elements.

A factorial set is said to be *right-extendable* if for every $w \in S$ there is some $a \in A$ such that $wa \in S$. It is *biextendable* if for any $w \in S$, there are $a, b \in A$ such that $awb \in S$.

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and if for every $u, w \in S$ there is a $v \in S$ such that $uvw \in S$. A recurrent set is biextendable. It is said to be *uniformly recurrent* if it is right-extendable and if, for any word $u \in S$, there exists an integer $n \geq 1$ such that $u$ is a factor of every word of $S$ of length $n$. A uniformly recurrent set is recurrent.

We denote by $A^{\mathbb{N}}$ the set of infinite words on the alphabet $A$. For a set $X \subset A^{\mathbb{N}}$, we denote by $F(X)$ the set of factors of the words of $X$.

Let $S$ be a set of words on the alphabet $A$. For $w \in S$, set $R(w) = \{a \in A \mid wa \in S\}$ and $L(w = \{a \in A \mid aw \in S\}$. A word $w$ is called *right-special* if $\mathrm{Card}(R(w)) \geq 2$ and *left-special* if $\mathrm{Card}(L(w)) \geq 2$. It is *bispecial* if it is both right and left-special.

An infinite word on a binary alphabet is *Sturmian* if its set of factors is closed under reversal and if for each $n$ there is exactly one right-special word of length $n$.

An infinite word is a *strict episturmian* word if its set of factors is closed under reversal and for each $n$ there is exactly one right-special word $w$ of length $n$, which is moreover such that $\mathrm{Card}(R(w)) = \mathrm{Card}(A)$.

A morphism $f : A^* \to A^*$ is called *primitive* if there is an integer $k$ such that for all $a, b \in A$, the letter $b$ appears in $f^k(a)$. If $f$ is a primitive morphism, the set of factors of any fixpoint of $f$ is uniformly recurrent (see [10], Proposition 1.2.3 for example).

**Example 2.5** Let $A = \{a, b\}$. The Fibonacci word is the fixpoint $x = f^\omega(a) = abaababa\ldots$ of the morphism $f : A^* \to A^*$ defined by $f(a) = ab$ and $f(b) = a$. It is a Sturmian word (see [13]). The set $F(x)$ of factors of $x$ is the *Fibonacci set.*

**Example 2.6** Let $A = \{a, b, c\}$. The Tribonacci word is the fixpoint $x = f^\omega(a) = abacaba\cdots$ of the morphism $f : A^* \to A^*$ defined by $f(a) = ab$, $f(b) = ac$, $f(c) = a$. It is a strict episturmian word (see [11]). The set $F(x)$ of factors of $x$ is the *Tribonacci set.*

## 2.4 Interval exchange sets

Let $T$ be an interval exchange set. The set $F(\Sigma_T(z))$ is called an *interval exchange set.* It is biextendable.

If $T$ is a minimal interval exchange transformation, one has $w \in F(\Sigma_T(z))$ if and only if $I_w \neq \emptyset$. Thus the set $F(\Sigma_T(z))$ does not depend on $z$. Since it depends only on $T$, we denote it by $F(T)$. When $T$ is regular (resp. minimal), such a set is called a *regular interval exchange set* (resp. a minimal interval exchange set).

Let $T$ be an interval exchange transformation. Let $M$ be the closure in $A^{\mathbb{N}}$ of the set of all $\Sigma_T(z)$ for $z \in [0, 1[$ and let $\sigma$ be the shift on $M$. The pair $(M, \sigma)$ is a *symbolic dynamical system*, formed of a topological space $M$ and a continuous transformation $\sigma$. Such a system is said to be minimal if the only closed subsets invariant by $\sigma$ are $\emptyset$ or $M$ (that is, every orbit is dense). It is well-known that $(M, \sigma)$ is minimal if and only if $F(T)$ is uniformly recurrent (see for example [13] Theorem 1.5.9).

We have the following commutative diagram (Figure 2.2).

$$\begin{tikzcd}
{[0,1[} \arrow[r, "T"] \arrow[d, "\Sigma_T"] & {[0,1[} \arrow[d, "\Sigma_T"] \\
M \arrow[r, "\sigma"] & M
\end{tikzcd}$$

Figure 2.2: The transformations $T$ and $\sigma$.

The map $\Sigma_T$ is neither continuous nor surjective. This can be corrected by embedding the interval $[0,1[$ into a larger space on which $T$ is a homeomophism (see [12] or [6] page 349). However, if the transformation $T$ is minimal, the symbolic dynamical system $(M, S)$ is minimal (see [6] page 392). Thus, we obtain the following statement.

**Proposition 2.7** *For any minimal interval exchange transformation $T$, the set $F(T)$ is uniformly recurrent.*

Note that for a minimal interval exchange transformation $T$, the map $\Sigma_T$ is injective (see [12] page 30).

The following is an elementary property of the intervals $I_u$ which will be used below. We denote by $<_1$ the lexicographic order on $A^*$ induced by the order $<_1$ on $A$.

**Proposition 2.8** *One has $I_u < I_v$ if and only if $u <_1 v$ and $u$ is not a prefix of $v$.*

*Proof.* For a word $u$ and a letter $a$, it results from (2.1) that $I_{ua} = I_u \cap T^{-|u|}(I_a)$. Since $(I_a)_{a\in A}$ is an ordered partition, this implies that $(T^{|u|}(I_u) \cap I_a)_{a\in A}$ is an ordered partition of $T^{|u|}(I_u)$. Since the restriction of $T^{|u|}$ to $I_u$ is a translation, this implies that $(I_{ua})_{a\in A}$ is an ordered partition of $I_u$. Moreover, for two words $u, v$, it results also from (2.1) that $I_{uv} = I_u \cap T^{-|u|}(I_v)$. Thus $I_{uv} \subset I_u$.

Assume that $u <_1 v$ and that $u$ is not a prefix of $v$. Then $u = \ell as$ and $v = \ell bt$ with $a, b$ two letters such that $a <_1 b$. Then we have $I_{\ell a} < I_{\ell b}$, with $I_u \subset I_{\ell a}$ and $I_v \subset I_{\ell b}$ whence $I_u < I_v$.

Conversely, assume that $I_u < I_v$. Since $I_u \cap I_v = \emptyset$, the words $u, v$ cannot be comparable for the prefix order. Set $u = \ell as$ and $v = \ell bt$ with $a, b$ two distinct letters. If $b <_1 a$, then $I_v < I_u$ as we have shown above. Thus $a <_1 b$ which implies $u <_1 v$. ∎

We denote by $<_2$ the order on $A^*$ defined by $u <_2 v$ if $u$ is a proper suffix of $v$ or if $u = waz$ and $v = tbz$ with $a <_2 b$. Thus $<_2$ is the lexicographic order on the reversal of the words induced by the order $<_2$ on the alphabet.

We denote by $\pi$ the morphism from $A^*$ onto itself which extends to $A^*$ the permutation $\pi$ on $A$. Then $u <_2 v$ if and only if $\pi^{-1}(\tilde{u}) <_1 \pi^{-1}(\tilde{v})$, where $\tilde{u}$ denotes the reversal of the word $u$.

The following statement is the analogue of Proposition 2.8.

**Proposition 2.9** *Let $T_{\lambda,\pi}$ be an interval exchange transformation. One has $J_u < J_v$ if and only if $u <_2 v$ and $u$ is not a suffix of $v$.*

*Proof.* Let $(I'_a)_{a\in A}$ be the family of semi-intervals defined by $I'_a = J_{\pi(a)}$. Then the interval exchange transformation $T'$ relative to $(I'_a)$ with translation values $-\alpha_a$ is the inverse of the transformation $T$. The semi-intervals $I'_w$ defined by Equation (2.1) with respect to $T'$ satisfy $I'_w = J_{\pi(\tilde{w})}$ or equivalently $J_w =$

$I'_{\pi^{-1}(\tilde{w})}$. Thus, $J_u < J_v$ if and only if $I'_{\pi^{-1}(\tilde{u})} < I'_{\pi^{-1}(\tilde{v})}$ if and only if (by Proposition 2.8) $\pi^{-1}(\tilde{u}) <_1 \pi^{-1}(\tilde{v})$ or equivalently $u <_2 v$. ■

# 3 Bifix codes and interval exchange

In this section, we first introduce prefix codes and bifix codes. For a more detailed exposition, see [3]. We describe the link between maximal bifix codes and interval exchange transformations and we prove our main result (Theorem 3.13).

## 3.1 Prefix codes and bifix codes

A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. A suffix code is defined symmetrically. A *bifix code* is a set which is both a prefix code and a suffix code.

A *coding morphism* for a prefix code $X \subset A^+$ is a morphism $f : B^* \to A^*$ which maps bijectively $B$ onto $X$.

Let $S$ be a set of words. A prefix code $X \subset S$ is $S$-maximal if it is not properly contained in any prefix code $Y \subset S$. Note that if $X \subset S$ is an $S$-maximal prefix code, any word of $S$ is comparable for the prefix order with a word of $X$.

A map $\lambda : A^* \to [0, 1]$ such that $\lambda(\varepsilon) = 1$ and, for any word $w$

$$\sum_{a \in A} \lambda(aw) = \sum_{a \in A} \lambda(wa) = \lambda(w), \tag{3.1}$$

is called an *invariant probability distribution* on $A^*$.

Let $T_{\lambda,\pi}$ be an interval exchange transformation. For any word $w \in A^*$, denote by $|I_w|$ the length of the semi-interval $I_w$ defined by Equation (2.1). Set $\lambda(w) = |I_w|$. Then $\lambda(\varepsilon) = 1$ and for any word $w$, Equation (3.1) holds and thus $\lambda$ is an invariant probability distribution.

The fact that $\lambda$ is an invariant probability measure is equivalent to the fact that the Lebesgue measure on $[0, 1[$ is invariant by $T$. It is known that almost all regular interval exchange transformations have no other invariant probability measure (and thus are uniquely ergodic, see [6] for references).

**Example 3.1** Let $S$ be the set of factors of the Fibonacci word (see Example 2.5). It is the natural coding of the rotation of angle $\alpha = (3 - \sqrt{5})/2$ with respect to $\alpha$ (see [13], Chapter 2). The values of the map $\lambda$ on the words of length at most 4 in $S$ are indicated in Figure 3.1.

The following result is a particular case of a result from [2] (Proposition 3.3.4).

**Proposition 3.2** *Let $T$ be a minimal interval exchange transformation, let $S = F(T)$ and let $\lambda$ be an invariant probability distribution on $S$. For any finite $S$-maximal prefix code $X$, one has $\sum_{x \in X} \lambda(x) = 1$.*

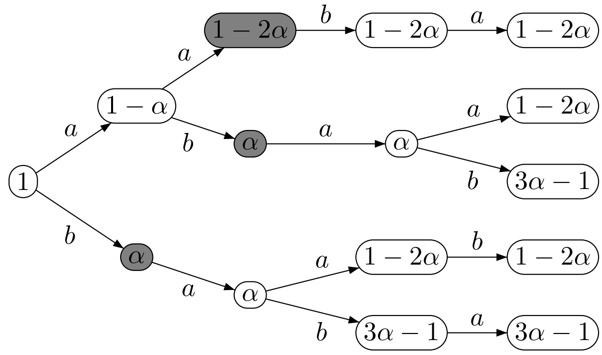


Figure 3.1: The invariant probability distribution on the Fibonacci set.

The following statement is connected with Proposition 3.2.

**Proposition 3.3** *Let $T$ be a minimal interval exchange transformation relative to $(I_a)_{a\in A}$, let $S = F(T)$ and let $X$ be a finite $S$-maximal prefix code ordered by $<_1$. The family $(I_w)_{w\in X}$ is an ordered partition of* $[0, 1[$.

*Proof.* By Proposition 2.8, the sets $(I_w)$ for $w \in X$ are pairwise disjoint. Let $\pi$ be the invariant probability distribution on $S$ defined by $\pi(w) = |I_w|$. By Proposition 3.2, we have $\sum_{w\in X} \pi(w) = 1$. Thus the family $(I_w)_{w\in X}$ is a partition of $[0, 1[$. By Proposition 2.8 it is an ordered partition. ∎

**Example 3.4** Let $T$ be the rotation of angle $\alpha = (3-\sqrt{5})/2$. The set $S = F(T)$ is the Fibonacci set. The set $X = \{aa, ab, b\}$ is an $S$-maximal prefix code (see the grey nodes in Figure 3.1). The partition of $[0, 1[$ corresponding to $X$ is

$$I_{aa} = [0, 1-2\alpha[, \quad I_{ab} = [1-2\alpha, 1-\alpha[, \quad I_b = [1-\alpha, 1[.$$

The values of the lengths of the semi-intervals (the invariant probability distribution) can also be read on Figure 3.1.

A symmetric statement holds for an $S$-maximal suffix code, namely that the family $(J_w)_{w\in X}$ is an ordered partition of $[0, 1[$ for the order $<_2$ on $X$.

## 3.2 Maximal bifix codes

Let $S$ be a set of words. A bifix code $X \subset S$ is $S$-maximal if it is not properly contained in a bifix code $Y \subset S$. For a recurrent set $S$, a finite bifix code is $S$-maximal as a bifix code if and only if it is an $S$-maximal prefix code (see [2], Theorem 4.2.2).

A *parse* of a word $w$ with respect to a bifix code $X$ is a triple $(v, x, u)$ such that $w = vxu$ where $v$ has no suffix in $X$, $u$ has no prefix in $X$ and $x \in X^*$. We denote by $\delta_X(w)$ the number of parses of $w$ with respect to $X$.

The number of parses of a word $w$ is also equal to the number of suffixes of $w$ which have no prefix in $X$ and the number of prefixes of $w$ which have no suffix in $X$ (see Proposition 6.1.6 in [3]).

By definition, the $S$-*degree* of a bifix code $X$, denoted $d_X(S)$, is the maximal number of parses of a word in $S$. It can be finite or infinite.

The set of *internal factors* of a set of words $X$, denoted $I(X)$, is the set of words $w$ such that there exist nonempty words $u, v$ with $uwv \in X$.

Let $S$ be a recurrent set and let $X$ be a finite $S$-maximal bifix code of $S$-degree $d$. A word $w \in S$ is such that $\delta_X(w) < d$ if and only if it is an internal factor of $X$, that is

$$I(X) = \{w \in S \mid \delta_X(w) < d\}$$

(Theorem 4.2.8 in [2]). Thus any word of $S$ which is not a factor of $X$ has $d$ parses. This implies that the $S$-degree $d$ is finite.

**Example 3.5** Let $S$ be a recurrent set. For any integer $n \geq 1$, the set $S \cap A^n$ is an $S$-maximal bifix code of $S$-degree $n$.

The *kernel* of a bifix code $X$ is the set $K(X) = I(X) \cap X$. Thus it is the set of words of $X$ which are also internal factors of $X$. By Theorem 4.3.11 of [2], a finite $S$-maximal bifix code is determined by its $S$-degree and its kernel.

**Example 3.6** Let $S$ be the Fibonacci set. The set $X = \{a, baab, bab\}$ is the unique $S$-maximal bifix code of $S$-degree 2 with kernel $\{a\}$. Indeed, the word $bab$ is not an internal factor and has two parses, namely $(1, bab, 1)$ and $(b, a, b)$.

The following result shows that bifix codes have a natural connection with interval exchange transformations.

**Proposition 3.7** *If $X$ is a finite $S$-maximal bifix code, with $S$ as in Proposition 3.3, the families $(I_w)_{w \in X}$ and $(J_w)_{w \in X}$ are ordered partitions of $[0, 1[$, relatively to the orders $<_1$ and $<_2$ respectively.*

*Proof.* This results from Proposition 3.3 and its symmetric and from the fact that, since $S$ is recurrent, a finite $S$-maximal bifix code is both an $S$-maximal prefix code and an $S$-maximal suffix code. ∎

Let $T$ be a regular interval exchange transformation relative to $(I_a)_{a \in A}$. Let $(\alpha_a)_{a \in A}$ be the translation values of $T$. Set $S = F(T)$. Let $X$ be a finite $S$-maximal bifix code on the alphabet $A$.

Let $T_X$ be the transformation on $[0, 1[$ defined by

$$T_X(z) = T^{|u|}(z) \quad \text{if} \quad z \in I_u$$

with $u \in X$. The transformation is well-defined since, by Proposition 3.7, the family $(I_u)_{u \in X}$ is a partition of $[0, 1[$.

Let $f : B^* \to A^*$ be a coding morphism for $X$. Let $(K_b)_{b \in B}$ be the family of semi-intervals indexed by the alphabet $B$ with $K_b = I_{f(b)}$. We consider $B$ as ordered by the orders $<_1$ and $<_2$ induced by $f$. Let $T_f$ be the interval exchange transformation relative to $(K_b)_{b\in B}$. Its translation values are $\beta_b = \sum_{j=0}^{m-1} \alpha_{a_j}$ for $f(b) = a_0 a_1 \cdots a_{m-1}$. The transformation $T_f$ is called the *transformation associated* with $f$.

**Proposition 3.8** *Let $T$ be a regular interval exchange transformation relative to $(I_a)_{a\in A}$ and let $S = F(T)$. If $f : B^* \to A^*$ is a coding morphism for a finite $S$-maximal bifix code $X$, one has $T_f = T_X$.*

*Proof.* By Proposition 3.7, the family $(K_b)_{b\in B}$ is a partition of $[0, 1[$ ordered by $<_1$. For any $w \in X$, we have by Equation (2.5) $J_w = I_w + \alpha_w$ and thus $T_X$ is the interval exchange transformation relative to $(K_b)_{b\in B}$ with translation values $\beta_b$. ■

In the sequel, under the hypotheses of Proposition 3.8, we consider $T_f$ as an interval exchange transformation. In particular, the natural coding of $T_f$ relative to $z \in [0, 1[$ is well-defined.

**Example 3.9** Let $S$ be the Fibonacci set. It is the set of factors of the Fibonacci word, which is a natural coding of the rotation of angle $\alpha = (3-\sqrt{5})/2$ relative to $\alpha$ (see Example 3.1). Let $X = \{aa, ab, ba\}$ and let $f$ be the coding morphism defined by $f(u) = aa$, $f(v) = ab$, $f(w) = ba$. The two partitions of $[0, 1[$ corresponding to $T_f$ are

$$I_u = [0, 1-2\alpha[, \quad I_v = [1-2\alpha, 1-\alpha[ \quad I_w = [1-\alpha, 1[$$

and

$$J_v = [0, \alpha[, \quad J_w = [\alpha, 2\alpha[ \quad J_u = [2\alpha, 1[$$

The transformation $T_f$ is represented in Figure 3.2. It is actually a representa-

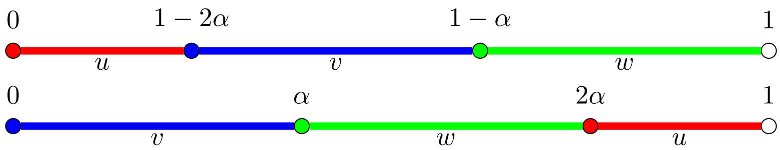


Figure 3.2: The transformation $T_f$.

tion on 3 intervals of the rotation of angle $2\alpha$. Note that the point $z = 1-\alpha$ is a separation point which is not a singularity of $T_f$. The first row of Table 3.1 gives the two orders on $X$. The next two rows give the two orders for each of the two other $S$-maximal bifix codes of $S$-degree 2 (there are actually exactly three $S$-maximal bifix codes of $S$-degree 2 in the Fibonacci set, see [2]).

Let $T$ be a minimal interval exchange transformation on the alphabet $A$. Let $x$ be the natural coding of $T$ relative to some $z \in [0, 1[$. Set $S = F(x)$. Let $X$ be a finite $S$-maximal bifix code. Let $f : B^* \to A^*$ be a morphism which maps bijectively $B$ onto $X$. Since $S$ is recurrent, the set $X$ is an $S$-maximal

| $(X, <_1)$ | $(X, <_2)$ |
|---|---|
| $aa, ab, ba$ | $ab, ba, aa$ |
| $a, baab, bab$ | $bab, baab, a$ |
| $aa, aba, b$ | $b, aba, aa$ |

Table 3.1: The two orders on the three $S$-maximal bifix codes of $S$-degree 2.

prefix code. Thus $x$ has a prefix $x_0 \in X$. Set $x = x_0x'$. In the same way $x'$ has a prefix $x_1$ in $X$. Iterating this argument, we see that $x = x_0x_1\cdots$ with $x_i \in X$. Consequently, there exists an infinite word $y$ on the alphabet $B$ such that $x = f(y)$. The word $y$ is the *decoding* of the infinite word $x$ with respect to $f$.

**Proposition 3.10** *The decoding of $x$ with respect to $f$ is the natural coding of the transformation associated with $f$ relative to $z$: $\Sigma_T(z) = f(\Sigma_{T_f}(z))$.*

*Proof.* Let $y = b_0b_1\cdots$ be the decoding of $x$ with respect to $f$. Set $x_i = f(b_i)$ for $i \geq 0$. Then, for any $n \geq 0$, we have

$$T_f^n(z) = T^{|u_n|}(z) \tag{3.2}$$

with $u_n = x_0\cdots x_{n-1}$ (note that $|u_n|$ denotes the length of $u_n$ with respect to the alphabet $A$). Indeed, this is is true for $n = 0$. Next $T_f^{n+1}(z) = T_f(t)$ with $t = T_f^n(z)$. Arguing by induction, we have $t = T^{|u_n|}(z)$. Since $x = u_nx_nx_{n+1}\cdots$, $t$ is in $I_{x_n}$ by (2.3). Thus by Proposition 3.8, $T_f(t) = T^{|x_n|}(t)$ and we obtain $T_f^{n+1}(z) = T^{|x_n|}(T^{|u_n|}(z)) = T^{|u_{n+1}|}(z)$ proving (3.2). Finally, for $u = f(b)$ with $b \in B$,

$$b_n = b \iff x_n = u \iff T^{|u_n|}(z) \in I_u \iff T_f^n(z) \in I_u = K_b$$

showing that $y$ is the natural coding of $T_f$ relative to $z$.

■

**Example 3.11** Let $T, \alpha, X$ and $f$ be as in Example 3.9. Let $x = abaababa\cdots$ be the Fibonacci word. We have $x = \Sigma_T(\alpha)$. The decoding of $x$ with respect to $f$ is $y = vuwwv\cdots$.

## 3.3 Bifix codes and regular transformations

The following result shows that for the coding morphism $f$ of a finite $S$-maximal bifix code, the map $T \mapsto T_f$ preserves the regularity of the transformation.

**Theorem 3.12** *Let $T$ be a regular interval exchange transformation and let $S = F(T)$. For any finite $S$-maximal bifix code $X$ with coding morphism $f$, the transformation $T_f$ is regular.*

*Proof.* Set $A = \{a_1, a_2, \ldots, a_s\}$ with $a_1 <_1 a_2 <_1 \cdots <_1 a_s$. We denote $\delta_i = \delta_{a_i}$. By hypothesis, the orbits of $\delta_2, \ldots, \delta_s$ are infinite and disjoint. Set $X = \{x_1, x_2, \ldots, x_t\}$ with $x_1 <_1 x_2 <_1 \cdots <_1 x_t$. Let $d$ be the $S$-degree of $X$.

For $x \in X$, denote by $\delta_x$ the left boundary of the semi-interval $J_x$. For each $x \in X$, it follows from Equation (2.2) that there is an $i \in \{1, \ldots, s\}$ such that $\delta_x = T^k(\delta_i)$ with $0 \leq k < |x|$. Moreover, we have $i = 1$ if and only if $x = x_1$. Since $T$ is regular, the index $i \neq 1$ and the integer $k$ are unique for each $x \neq x_1$. And for such $x$ and $i$, by (2.4), we have $\Sigma_T(\delta_i) = u\Sigma_T(\delta_x)$ with $u$ a proper suffix of $x$.

We now show that the orbits of $\delta_{x_2}, \ldots, \delta_{x_t}$ for the transformation $T_f$ are infinite and disjoint. Assume that $\delta_{x_p} = T_f^n(\delta_{x_q})$ for some $p, q \in \{2, \ldots, t\}$ and $n \in \mathbb{Z}$. Interchanging $p, q$ if necessary, we may assume that $n \geq 0$. Let $i, j \in \{2, \ldots, s\}$ be such that $\delta_{x_p} = T^k(\delta_i)$ with $0 \leq k < |x_p|$ and $\delta_{x_q} = T^\ell(\delta_j)$ with $0 \leq \ell < |x_q|$. Since $T^k(\delta_i) = T_f^n(T^\ell(\delta_j)) = T^{m+\ell}(\delta_j)$ for some $m \geq 0$, we cannot have $i \neq j$ since otherwise the orbits of $\delta_i, \delta_j$ for the transformation $T$ intersect. Thus $i = j$. Since $\delta_{x_p} = T^k(\delta_i)$, we have $\Sigma_T(\delta_i) = u\Sigma_T(\delta_{x_p})$ with $|u| = k$, $u$ proper suffix of $x_p$. And since $\delta_{x_p} = T_f^n(\delta_{x_q})$, we have $\Sigma_T(\delta_{x_q}) = x\Sigma_T(\delta_{x_p})$ with $x \in X^*$. Since on the other hand $\delta_{x_q} = T^\ell(\delta_i)$, we have $\Sigma_T(\delta_i) = v\Sigma_T(\delta_{x_q})$ with $|v| = \ell$ and $v$ a proper suffix of $x_q$. We obtain

$$\begin{aligned}\Sigma_T(\delta_i) &= u\Sigma_T(\delta_{x_p}) \\ &= v\Sigma_T(\delta_{x_q}) = vx\Sigma_T(\delta_{x_p})\end{aligned}$$

Since $|u| = |vx|$, this implies $u = vx$. But since $u$ cannot have a suffix in $X$, $u = vx$ implies $x = 1$ and thus $n = 0$ and $p = q$. This concludes the proof. ∎

Let $f$ be a coding morphism for a finite $S$-maximal bifix code $X \subset S$. The set $f^{-1}(S)$ is called a *maximal bifix decoding* of $S$.

**Theorem 3.13** *The family of regular interval exchange sets is closed under maximal bifix decoding.*

*Proof.* Let $T$ be a regular interval exchange transformation such that $S = F(T)$. By Theorem 3.12, $T_f$ is a regular interval exchange transformation. We show that $f^{-1}(S) = F(T_f)$, which implies the conclusion.

Let $x = \Sigma_T(z)$ for some $z \in [0, 1[$ and let $y = f^{-1}(x)$. Then $S = F(x)$ and $F(T_f) = F(y)$. For any $w \in F(y)$, we have $f(w) \in F(x)$ and thus $w \in f^{-1}(S)$. This shows that $F(T_f) \subset f^{-1}(S)$. Conversely, let $w \in f^{-1}(S)$ and let $v = f(w)$. Since $S = F(x)$, there is a word $u$ such that $uv$ is a prefix of $x$. Set $z' = T^{|u|}(z)$ and $x' = \Sigma_T(z')$. Then $v$ is a prefix of $x'$ and $w$ is a prefix of $y' = f^{-1}(x')$. Since $T_f$ is regular, it is minimal and thus $F(y') = F(T_f)$. This implies that $w \in F(T_f)$. ∎

Since a regular interval exchange set is uniformly recurrent, Theorem 3.13 implies in particular that if $S$ is a regular interval exchange set and $f$ a coding morphism of a finite $S$-maximal bifix code, then $f^{-1}(S)$ is uniformly recurrent.

This is not true for an arbitrary uniformly recurrent set $S$, as shown by the following example.

**Example 3.14** Set $A = \{a, b\}$ and $B = \{u, v\}$. Let $S$ be the set of factors of $(ab)^*$ and let $f : B^* \to A^*$ be defined by $f(u) = ab$ and $f(v) = ba$. Then $f^{-1}(S) = u^* \cup v^*$ which is not recurrent.

We illustrate the proof of Theorem 3.12 in the following example.

**Example 3.15** Let $T$ be the rotation of angle $\alpha = (3 - \sqrt{5})/2$. The set $S = F(T)$ is the Fibonacci set. Let $X = \{a, baab, babaabaabab, babaabab\}$. The set $X$ is an $S$-maximal bifix code of $S$-degree 3 (see [2]). The values of the $\mu_{x_i}$ (which are the right boundaries of the intervals $I_{x_i}$) and $\delta_{x_i}$ are represented in Figure 3.3.

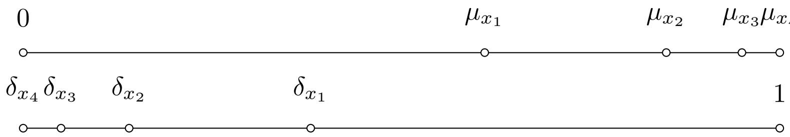


Figure 3.3: The transformation associated with a bifix code of $S$-degree 3.

The infinite word $\Sigma_T(0)$ is represented in Figure 3.4. The value indicated on the word $\Sigma_T(0)$ after a prefix $u$ is $T^{|u|}(0)$. The three values $\delta_{x_4}, \delta_{x_2}, \delta_{x_3}$ correspond to the three prefixes of $\Sigma_T(0)$ which are proper suffixes of $X$.

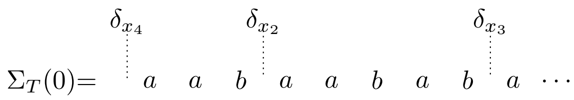


Figure 3.4: The infinite word $\Sigma_T(0)$.

The following example shows that Theorem 3.13 is not true when $X$ is not bifix.

**Example 3.16** Let $S$ be the Fibonacci set and let $X = \{aa, ab, b\}$. The set $X$ is an $S$-maximal prefix code. Let $B = \{u, v, w\}$ and let $f$ be the coding morphism for $X$ defined by $f(u) = aa$, $f(v) = ab$, $f(w) = b$. The set $W = f^{-1}(S)$ is not an interval exchange set. Indeed, we have $vu, vv, wu, wv \in W$. This implies that both $J_v$ and $J_w$ meet $I_u$ and $I_v$, which is impossible in an interval exchange transformation.

# 4 Tree sets

We introduce in this section the notions of tree sets and planar tree sets. We first introduce the notion of extension graph which describes the possible two-sided extensions of a word.

## 4.1 Extension graphs

Let $S$ be a biextendable set of words. For $w \in S$, we denote

$$L(w) = \{a \in A \mid aw \in S\}, \quad R(w) = \{a \in A \mid wa \in S\}$$

and

$$E(w) = \{(a,b) \in A \times A \mid awb \in S\}.$$

For $w \in S$, the *extension graph* of $w$ is the undirected bipartite graph $G(w)$ on the set of vertices which is the disjoint union of two copies of $L(w)$ and $R(w)$ with edges the pairs $(a,b) \in E(w)$.

Recall that an undirected graph is a tree if it is connected and acyclic.

Let $S$ be a biextendable set. We say that $S$ is a *tree set* if the graph $G(w)$ is a tree for all $w \in S$.

Let $<_1$ and $<_2$ be two orders on $A$. For a set $S$ and a word $w \in S$, we say that the graph $G(w)$ is *compatible* with the orders $<_1$ and $<_2$ if for any $(a,b),(c,d) \in E(w)$, one has

$$a <_1 c \Longrightarrow b \leq_2 d.$$

Thus, placing the vertices of $L(w)$ ordered by $<_1$ on a line and those of $R(w)$ ordered by $<_2$ on a parallel line, the edges of the graph may be drawn as straight noncrossing segments, resulting in a planar graph.

We say that a biextendable set $S$ is a *planar tree set* with respect to two orders $<_1$ and $<_2$ on $A$ if for any $w \in S$, the graph $G(w)$ is a tree compatible with $<_1, <_2$. Obviously, a planar tree set is a tree set.

The following example shows that the Tribonacci set is not a planar tree set.

**Example 4.1** Let $S$ be the Tribonacci set (see example 2.6). The words $a, aba$ and $abacaba$ are bispecial. Thus the words $ba, caba$ are right-special and the words $ab, abac$ are left-special. The graphs $G(\varepsilon), G(a)$ and $G(aba)$ are shown in Figure 4.1. One sees easily that it not possible to find two orders on $A$ making

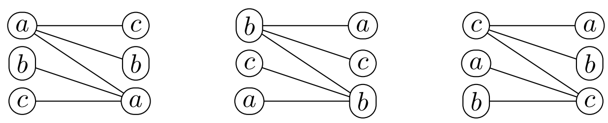


Figure 4.1: The graphs $G(\varepsilon), G(a)$ and $G(aba)$ in the Tribonacci set.

the three graphs planar.

## 4.2 Interval exchange sets and planar tree sets

The following result is proved in [9] with a converse (see below).

**Proposition 4.2** *Let $T$ be an interval exchange transformation on $A$ ordered by $<_1$ and $<_2$. If $T$ is regular, the set $F(T)$ is a planar tree set with respect to $<_2$ and $<_1$.*

*Proof.* Assume that $T$ is a regular interval exchange transformation relative to $(I_a, \alpha_a)_{a\in A}$ and let $S = F(T)$.

Since $T$ is minimal, $w$ is in $S$ if and only if $I_w \neq \emptyset$. Thus, one has

(i) $b \in R(w)$ if and only if $I_w \cap T^{-|w|}(I_b) \neq \emptyset$ and

(ii) $a \in L(w)$ if and only if $J_a \cap I_w \neq \emptyset$.

Condition (i) holds because $I_{wb} = I_w \cap T^{-|w|}(I_b)$ and condition (ii) because $I_{aw} = I_a \cap T^{-1}(I_w)$, which implies $T(I_{aw}) = J_a \cap I_w$. In particular, (i) implies that $(I_{wb})_{b\in R(w)}$ is an ordered partition of $I_w$ with respect to $<_1$.

We say that a path in a graph is reduced if does not use consecutively the same edge. For $a, a' \in L(w)$ with $a <_2 a'$, there is a unique reduced path in $G(w)$ from $a$ to $a'$ which is the sequence $a_1, b_1, \ldots a_n$ with $a_1 = a$ and $a_n = a'$ with $a_1 <_2 a_2 <_2 \cdots <_2 a_n$, $b_1 <_1 b_2 <_1 \cdots <_1 b_{n-1}$ and $J_{a_i} \cap I_{wb_i} \neq \emptyset$, $J_{a_{i+1}} \cap I_{wb_i} \neq \emptyset$ for $1 \leq i \leq n-1$ (see Figure 4.2). Note that the hypothesis that $T$ is regular is needed here since otherwise the right boundary of $J_{a_i}$ could be the left boundary of $I_{wb_i}$. Thus $G(w)$ is a tree. It is compatible with $<_2, <_1$ since the above shows that $a <_2 a'$ implies that the letters $b_1, b_{n-1}$ such that $(a, b_1), (a', b_{n-1}) \in E(w)$ satisfy $b_1 \leq_1 b_{n-1}$. ■

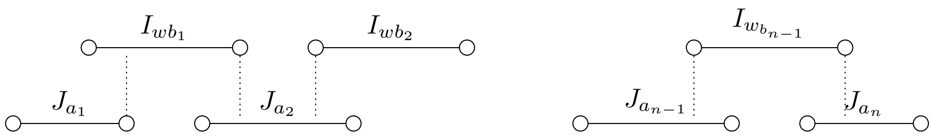


Figure 4.2: A path from $a_1$ to $a_n$ in $G(w)$.

By Proposition 4.2, a regular interval exchange set is a planar tree set, and thus in particular a tree set. Note that the analogue of Theorem 3.13 holds for the class of uniformly recurrent tree sets [4].

The main result of [9] states that a uniformly recurrent set $S$ on an alphabet $A$ is a regular interval exchange set if and only if $A \subset S$ and there exist two orders $<_1$ and $<_2$ on $A$ such that the following conditions are satisfied for any word $w \in S$.

(i) The set $L(w)$ (resp. $R(w)$) is formed of consecutive elements for the order $<_2$ (resp. $<_1$).

(ii) For $(a, b), (c, d) \in E(w)$, if $a <_2 c$, then $b \leq_1 d$.

(iii) If $a, b \in L(w)$ are consecutive for the order $<_2$, then the set $R(aw) \cap R(bw)$ is a singleton.

It is easy to see that a biextendable set $S$ containing $A$ satisfies (ii) and (iii) if and only if it is a planar tree set. Actually, in this case, it automatically satisfies also condition (i). Indeed, let us consider a word $w$ and $a, b, c \in A$ with $a <_1 b <_1 c$ such that $wa, wc \in S$ but $wb \notin S$. Since $b \in S$ there is a (possibly empty) suffix $v$ of $w$ such that $vb \in S$. We choose $v$ of maximal length. Since $wb \notin S$, we have $w = uv$ with $u$ nonempty. Let $d$ be the last letter of $u$. Then we have $dva, dvc \in S$ and $dvb \notin S$. Since $G(v)$ is a tree and $b \in R(v)$, there is a letter $e \in L(v)$ such that $evb \in S$. But $e <_2 d$ and $d <_2 e$ are both impossible since $G(v)$ is compatible with $<_2$ and $<_1$. Thus we reach a contradiction.

This shows that the following reformulation of the main result of [9] is equivalent to the original one.

**Theorem 4.3 (Ferenczi, Zamboni)** *A set $S$ is a regular interval exchange set on the alphabet $A$ if and only if it is a uniformly recurrent planar tree set containing $A$.*

We have already seen that the Tribonacci set is a tree set which is not a planar tree set (Example 4.1). The next example shows that there are uniformly recurrent tree sets which are neither Sturmian nor regular interval exchange sets.

**Example 4.4** Let $S$ be the Tribonacci set on the alphabet $A = \{a, b, c\}$ and let $f : \{x, y, z, t, u\}^* \to A^*$ be the coding morphism for $X = S \cap A^2$ defined by $f(x) = aa$, $f(y) = ab$, $f(z) = ac$, $f(t) = ba$, $f(u) = ca$. By Theorem 7.1 in [4], the set $W = f^{-1}(S)$ is a uniformly recurrent tree set. It is not Sturmian since $y$ and $t$ are two right-special words. It is not either a regular interval exchange set. Indeed, for any right-special word $w$ of $W$, one has $\mathrm{Card}(R(w)) = 3$. This is not possible in a regular interval exchange set $T$ since, $\Sigma_T$ being injective, the length of the interval $J_w$ tends to 0 as $|w|$ tends to infinity and it cannot contain several separation points. It can of course also be verified directly that $W$ is not a planar tree set.

### 4.3 Exchange of pieces

In this section, we show how one can define a generalization of interval exchange transformations called exchange of pieces. In the same way as interval exchange is a generalization of rotations on the circle, exchange of pieces is a generalization of rotations of the torus. We begin by studying this direction starting from the Tribonacci word. For more on the Tribonacci word, see [17] and also [14, Chap. 10].

**The Tribonacci shift** The Tribonacci set $S$ is not an interval exchange set but it is however the natural coding of another type of geometric transformation, namely an exchange of pieces in the plane, which is also a translation acting on the two-dimensional torus $\mathbb{T}^2$. This will allow us to show that the decoding of the Tribonacci word with respect to a coding morphism for a finite $S$-maximal bifix code is again a natural coding of an exchange of pieces.

The *Tribonacci shift* is the symbolic dynamical system $(M_x, \sigma)$, where $M_x = \overline{\{\sigma^n(x) : n \in \mathbb{N}\}}$ is the closure of the $\sigma$-orbit of $x$ where $x$ is the Tribonacci word. By uniform recurrence of the Tribonacci word, $(M_x, \sigma)$ is minimal and $M_x = M_y$ for each $y \in M_x$ ([16, Proposition 4.7]). The Tribonacci set is the set of factors of the Tribonacci shift $(M_x, \sigma)$.

**Natural coding** Let $\Lambda$ be a full-rank lattice in $\mathbb{R}^d$. We say that an infinite word $x$ is a *natural coding* of a toral translation $T_{\mathbf{t}} : \mathbb{R}^d/\Lambda \to \mathbb{R}^d/\Lambda,\ \mathbf{x} \mapsto \mathbf{x} + \mathbf{t}$ if there exists a fundamental domain $R$ for $\Lambda$ together with a partition $R = R_1 \cup \cdots \cup R_k$ such that on each $R_i$ $(1 \le i \le k)$, there exists a vector $\mathbf{t}_i$ such that the map $T_{\mathbf{t}}$ is given by the translation along $\mathbf{t}_i$, and $x$ is the coding of a point $\mathbf{x} \in R$ with respect to this partition. A symbolic dynamical system $(M, \sigma)$ is a *natural coding* of $(\mathbb{R}^d/\Lambda, T_{\mathbf{t}})$ if every element of $M$ is a natural coding of the orbit of some point of the $d$-dimensional torus $\mathbb{R}^d/\Lambda$ (with respect to the same partition) and if, furthermore, $(M, \sigma)$ and $(\mathbb{R}^d/\Lambda, T_{\mathbf{t}})$ are measurably conjugate.

**Definition of the Rauzy fractal** Let $\beta$ stand for the Perron-Frobenius eigenvalue of the Tribonacci substitution. It is the largest root of $z^3 - z^2 - z - 1$. Consider the translation $R_\beta \colon \mathbb{T}^2 \to \mathbb{T}^2$, $x \mapsto x + (1/\beta, 1/\beta^2)$. Rauzy introduces in [18] a fundamental domain for a two-dimensional lattice, called the Rauzy fractal (it has indeed fractal boundary), which provides a partition for the symbolic dynamical system $(M_x, \sigma)$ to be a natural coding for $R_\beta$. The Tribonacci word is a natural coding of the orbit of the point 0 under the action of the toral translation in $\mathbb{T}^2$: $x \mapsto x + (\frac{1}{\beta}, \frac{1}{\beta^2})$. Similarly as in the case of interval exchanges, we have the following commutative diagram

$$\begin{array}{ccc} \mathbb{T}^2 & \xrightarrow[R_\beta]{} & \mathbb{T}^2 \\ \downarrow & & \downarrow \\ M_x & \xrightarrow{\sigma} & M_x \end{array}$$

The *Abelianization map* $\mathbf{f}$ of the free monoid $\{1,2,3\}^*$ is defined by $\mathbf{f} : \{1,2,3\}^* \to \mathbb{Z}^3$, $\mathbf{f}(w) = |w|_1 e_1 + |w|_2 e_2 + |w|_3 e_3$, where $|w|_i$ denotes the number of occurrences of the letter $i$ in the word $w$, and $(e_1, e_2, e_3)$ stands for the canonical basis of $\mathbb{R}^3$.

Let $f$ be the morphism $a \mapsto ab, b \mapsto ac, c \mapsto a$ such that the Tribonacci word is the fixpoint of $f$ (see Example 2.6). The incidence matrix $F$ of $f$ is defined by $F = (|f(j)|_i)_{(i,j) \in \mathcal{A}^2}$, where $|f(j)|_i$ counts the number of occurrences of the letter $i$ in $f(j)$. One has $F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. The incidence matrix $F$ admits as eigenspaces in $\mathbb{R}^3$ one *expanding eigenline* (generated by the eigenvector with positive coordinates $v_\beta = (1/\beta, 1/\beta^2, 1/\beta^3)$ associated with the eigenvalue $\beta$).

We consider the projection $\pi$ onto the antidiagonal plane $x+y+z=0$ along the expanding direction of the matrix $F$.

One associates with the Tribonacci word $x=(x_n)_{n\geq 0}$ a broken line starting from 0 in $\mathbb{Z}^3$ and approximating the expanding line $v_\beta$ as follows. The *Tribonacci broken line* is defined as the broken line which joins with segments of length 1 the points $\mathbf{f}(x_0x_1\cdots x_{n-1})$, $n\in\mathbb{N}$. In other words we describe this broken line by starting from the origin, and then by reading successively the letters of the Tribonacci word $x$, going one step in direction $e_i$ if one reads the letter $i$. The vectors $\mathbf{f}(x_0x_1\cdots x_n)$, $n\in\mathbb{N}$, stay within bounded distance of the expanding line (this comes from the fact that $\beta$ is a Pisot number). The closure of the set of projected vertices of the broken line is called the *Rauzy fractal* and is represented on Figure 4.3. We thus define the Rauzy fractal $\mathcal{R}$ as

$$\mathcal{R} := \overline{\{\pi(\mathbf{f}(x_0\cdots x_{n-1}));\ n\in\mathbb{N}\}},$$

where $x_0\ldots x_{n-1}$ stands for the empty word when $n=0$.

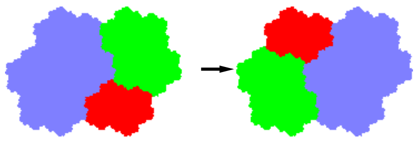

Figure 4.3: The Rauzy fractal

The Rauzy fractal is divided into three pieces, for $i=\{1,2,3\}$

$$\begin{aligned}\mathcal{R}(i) &:= \overline{\{\pi(\mathbf{f}(x_0\cdots x_{n-1}));\ x_n=i, n\in\mathbb{N}\}},\\ \mathcal{R}'(i) &:= \overline{\{\pi(\mathbf{f}(x_0\cdots x_n));\ x_n=i, n\in\mathbb{N}\}}.\end{aligned}$$

It has been proved in [18] that these pieces have non-empty interior and are disjoint up to a set of zero measure. The following exchange of pieces $E$ is thus well-defined

$$E:\mathrm{Int}\ \mathcal{R}_1\cup\mathrm{Int}\ \mathcal{R}_2\cup\mathrm{Int}\ \mathcal{R}_3\to\mathcal{R},\ x\mapsto x+\pi(e_i),\ \text{when } x\in\mathrm{Int}\ \mathcal{R}_i.$$

One has $E(\mathcal{R}_i)=\mathcal{R}_i'$, for all $i$.

We consider the lattice $\Lambda$ generated by the vectors $\pi(e_i)-\pi(e_j)$, for $i\neq j$. The Rauzy fractal tiles periodically the plane, that is, $\cup_{\gamma\in\Lambda}\gamma+\mathcal{R}$ is equal to the plane $x+y+z=0$, and for $\gamma\neq\gamma'\in\Lambda$, $\gamma+\mathcal{R}$ and $\gamma'+\mathcal{R}$ do not intersect (except on a set of zero measure). This is why the exchange of pieces is in fact measurably conjugate to the translation $R_\beta$. Indeed the vector of coordinates of $\pi(\mathbf{f}(x_0x_1\cdots x_{n-1}))$ in the basis $(\pi(e_3)-\pi(e_1),\pi(e_3)-\pi(e_2))$ of the plane $x+y+z=0$ is $n\cdot(1/\beta,1/\beta^2)-(|x_0x_1\cdots x_{n-1}|_1,|x_0x_1\cdots x_{n-1}|_2)$. Hence the coordinates of $E^n(0)$ in the basis $(e_3-e_1,e_3-e_2)$ are equal to $R_\beta^n(0)$ modulo $\mathbb{Z}^2$.

**Bifix codes and exchange of pieces** Let $(\mathcal{R}_a)_{a\in A}$ and $(\mathcal{R}'_a)_{a\in A}$ be two families of subsets of a compact set $\mathcal{R}$ incuded in $\mathbb{R}^d$. We assume that the families $(\mathcal{R}_a)_{a\in A}$ and the $(\mathcal{R}'_a)_{a\in A}$ both form a partition of $\mathcal{R}$ up to a set of zero measure. We assume that there exist vectors $e_a$ such that $\mathcal{R}'_a = \mathcal{R}_a + e_a$ for any $a \in A$. The exchange of pieces associated with these data is the map $E$ defined on $\mathcal{R}$ (except a set of measure zero) by $E(z) = z + e_a$ if $z \in \mathcal{R}_a$. The notion of natural coding of an exchange of pieces extends here in a natural way.

Assume that $E$ is an exchange of pieces as defined above. Let $S$ be the set of factors of the natural codings of $E$. We assume that $S$ is uniformly recurrent.

By analogy with the case of interval exchanges, let $I_a = \mathcal{R}_a$ and let $J_a = E(\mathcal{R}_a)$. For a word $w \in A^*$, one defines similarly as for interval exchanges $I_w$ and $J_w$.

Let $X$ be a finite $S$-maximal prefix code. The family $I_w$, $w \in X$, is a partition (up to sets of zero measure) of $\mathcal{R}$. If $X$ is a finite $S$-maximal suffix code, then the family $J_w$ is a partition (up to sets of zero measure) of $\mathcal{R}$. Let $f$ be a coding morphism for $X$. If $X$ is a finite $S$-maximal bifix code, then $E_X$ is the exchange of pieces $E_f$ (defined as for interval exchanges), hence the decoding of $x$ with respect to $f$ is the natural coding of the exchange of pieces associated with $f$. In particular, $S$ being the Tribonacci set, the decoding of $S$ by a finite $S$-maximal bifix code is again the natural coding of an exchange of pieces. If $X$ is the set of factors of length $n$ of $S$, then $E_f$ is in fact equal to $R_\beta^n$ (otherwise, there is no reason for this exchange of pieces to be a translation). The analogues of Proposition 3.8 and 3.10 thus hold here also.

### 4.4 Subgroups of finite index

We denote by $FG(A)$ the free group on the set $A$.

Let $S$ be a recurrent set containing the alphabet $A$. We say that $S$ has the *finite index basis property* if the following holds: a finite bifix code $X \subset S$ is an $S$-maximal bifix code of $S$-degree $d$ if and only if it is a basis of a subgroup of index $d$ of $FG(A)$.

The following is a consequence of the main result of [5].

**Theorem 4.5** *A regular interval exchange set has the finite index basis property.*

*Proof.* Let $T$ be a regular interval exchange transformation and let $S = F(T)$. Since $T$ is regular, $S$ is uniformly recurrent and by Proposition 4.2, it is a tree set. By Theorem 4.4 in [5], a uniformly recurrent tree set has the finite index basis property, and thus the conclusion follows. ■

Note that Theorem 4.5 implies in particular that if $T$ is a regular $s$-interval exchange set and if $X$ is a finite $S$-maximal bifix code of $S$-degree $d$, then $\mathrm{Card}(X) = d(s-1)+1$. Indeed, by Schreier's Formula a basis of a subgroup of index $d$ in a free group of rank $s$ has $d(s-1)+1$ elements.

We use Theorem 4.5 to give another proof of Theorem 3.12. For this, we recall the following notion.

Let $T$ be an interval exchange transformation on $I = [0,1[$ relative to $(I_a)_{a\in A}$. Let $G$ be a transitive permutation group on a finite set $Q$. Let $\varphi : A^* \to G$ be a morphism and let $\psi$ be the map from $I$ into $G$ defined by $\psi(z) = \varphi(a)$ if $z \in I_a$. The *skew product* of $T$ and $G$ is the transformation $U$ on $I \times Q$ defined by

$$U(z,q) = (T(z), q\psi(z))$$

(where $q\psi(z)$ is the result of the action of the permutation $\psi(z)$ on $q \in Q$). Such a transformation is equivalent to an interval exchange transformation via the identification of $I \times Q$ with an interval obtained by placing the $d = \mathrm{Card}(Q)$ copies of $I$ in sequence. This is called an *interval exchange transformation on a stack* in [7] (see also [19]). If $T$ is regular, then $U$ is also regular.

Let $T$ be a regular interval exchange transformation and let $S = F(T)$. Let $X$ be a finite $S$-maximal bifix code of $S$-degree $d = d_X(S)$. By Theorem 4.5, $X$ is a basis of a subgroup $H$ of index $d$ of $FG(A)$. Let $G$ be the representation of $FG(A)$ on the right cosets of $H$ and let $\varphi$ be the natural morphism from $FG(A)$ onto $G$. We identify the right cosets of $H$ with the set $Q = \{1, 2, \ldots, d\}$ with 1 identified to $H$. Thus $G$ is a transitive permutation group on $Q$ and $H$ is the inverse image by $\varphi$ of the permutations fixing 1.

The transformation induced by the skew product $U$ on $I \times \{1\}$ is clearly equivalent to the transformation $T_f = T_X$ where $f$ is a coding morphism for the $S$-maximal bifix code $X$. Thus $T_X$ is a regular interval exchange transformation.

**Example 4.6** Let $T$ be the rotation of Example 3.1. Let $Q = \{1,2,3\}$ and let $\varphi$ be the morphism from $A^*$ into the symmetric group on $Q$ defined by $\varphi(a) = (23)$ and $\varphi(b) = (12)$. The transformation induced by the skew product of $T$ and $G$ on $I \times \{1\}$ corresponds to the bifix code $X$ of Example 3.15. For example, we have $U : (1-\alpha, 1) \to (0,2) \to (\alpha,3) \to (2\alpha,2) \to (3\alpha-1,1)$ (see Figure 4.4) and the corresponding word of $X$ is $baab$.

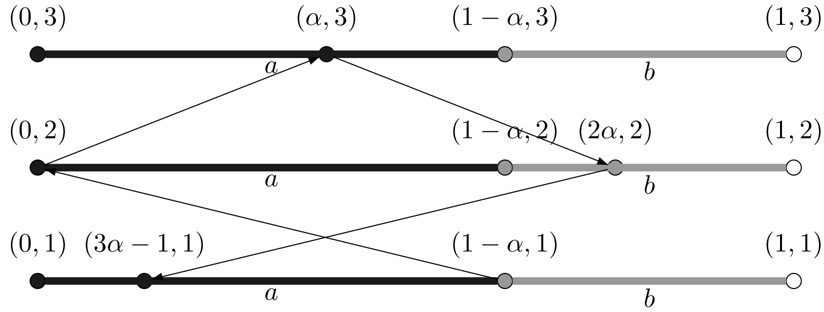


Figure 4.4: The transformation $U$.

# References

[1] Vladimir I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat. Nauk*, 18(6 (114)):91–192, 1963. 2

[2] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012. 2, 9, 10, 11, 12, 15

[3] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009. 5, 9, 11

[4] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. 2013. `http://arxiv.org/abs/1308.5396`. 2, 17, 18

[5] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. 2014. `http://arxiv.org/abs/1305.0127`. 2, 21

[6] Valérie Berthé and Michel Rigo, editors. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010. 3, 8, 9

[7] Michael D. Boshernitzan and C. R. Carroll. An extension of Lagrange's theorem to interval exchange transformations over quadratic fields. *J. Anal. Math.*, 72:21–44, 1997. 2, 22

[8] Isaac P. Cornfeld, Sergei V. Fomin, and Yakov G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ. 3

[9] Sébastien Ferenczi and Luca Q. Zamboni. Languages of $k$-interval exchange transformations. *Bull. Lond. Math. Soc.*, 40(4):705–714, 2008. 2, 17, 18

[10] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel. 7

[11] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000. 7

[12] Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975. 2, 5, 8

[13] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002. 7, 9

[14] M. Lothaire. *Applied combinatorics on words*, volume 105 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005. 18

[15] V. I. Oseledec. The spectrum of ergodic automorphisms. *Dokl. Akad. Nauk SSSR*, 168:1009–1011, 1966. 2

[16] Martine Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2010. 19

[17] Gérard Rauzy. Échanges d'intervalles et transformations induites. *Acta Arith.*, 34(4):315–328, 1979. 4, 18

[18] Gérard Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. France*, 110(2):147–178, 1982. 19, 20

[19] William A. Veech. Finite group extensions of irrational rotations. *Israel J. Math.*, 21(2-3):240–259, 1975. Conference on Ergodic Theory and Topological Dynamics (Kibbutz Lavi, 1974). 2, 22